ANALYSING THE SCORE MATCHING OF DNA SEQUENCING USING AN EXPERT SYSTEM OF NEURO-FUZZY

[1]SAFA A. HAMEED, [2]RAED I. HAMED

[1]College of computer science and Information technology, University of Anbar, Department of CS, Al-Anbar, Iraq

[2]College of Science and Technology, University of Human Development, Department of CS, Sulaymaniyah, Iraq

[1]la_programmer89@yahoo.com, [2]raed.alfalahy@uhd.edu.iq

ABSTRACT

The proposed approach in obtaining the matching score of the DNA sequence alignment is presented, the Neuro-Fuzzy technique is the method which are implemented in this paper, it's utilized the set of biological data of the DNA sequence performing with Global and Local algorithms in order to optimize the optimal alignment, we use the pairwise DNA sequence alignment to measure the score of the similarity, which depends on the scoring matrix to guarantee the confidence measure score of the pairwise sequence alignment, the ANFIS model is suitable for predicting the matching score through the training and testing in Neural network and the inference fuzzy system in fuzzy logic, that achieves the result in high performance implementation.

Keywords: *Sequence Alignment; Dynamic Programming; Matching Mechanism; Neuro-Fuzzy.*
Abbreviations: *DNA: Deoxyribo Nucleic Acid; ANFIS: Adaptive Neuro Fuzzy Inference System; GA: Global alignment; LA: Local Alignment ; MF: Membership Function.*

1. INTRODUCTION

This DNA sequence matching is an essential area and more approaching nearby in computational biological data [1]. DNA sequence analysis is an imperative exploration topic in Bioinformatics [2]. Presently Bioinformatics and genomics cover an extensive variety of information sites that are put away, utilized, and controlled by researchers and machines [3].

A noteworthy test of displaying biological systems is that customary strategies in light of physical and compound standards require information that is hard to precisely and reliably acquire utilizing either conventional biochemical or high throughput advances [4]. Finding and comprehension the complex causal connections inside gene through analysis technique [5]. In any case, it has been recognized the data contained in DNA sequences are difficult for people to understand without cautious extraction and preparing [6]. The latest mechanical advances have essentially diminished the expense of DNA sequencing [7]. The DNA atom contains biological, physical, and chemical data, it has turned out to be essential to examine DNA sequences statistically [8]. String matching is a strategy to find a design from the predefined info string [9]. DNA sequences advance by local changes influencing one or a few adjacent symbols and in addition by large scale improvements and duplications [10]. Sequence alignment is a discretionary matching process, thus there is a need for better algorithms [11]. DNA sequence alignment algorithms over computational biological science have been enhanced eventually by different techniques [12].

The Dynamic programming is the method to implement the DNA alignment using the Needleman-Wunsch [13] and Smith-Waterman [14]. The (Needleman-Wunsch) global, the (Smith-Waterman) local, and (ends-free) cover pairwise sequence alignment issues [15]. Pairwise Alignment a technique for scoring the similarity of a pair of Characters. It decides the correspondences between the substrings in the sequences like the similarity score is amplified [20]. For it's a large portion basic form, known as pairwise sequence alignment, we provided for two sequences A and B and discover their best alignment (either global or local) [21]. In the two sequences have a similar base, they are defined to be homologous [22]. Similarities between DNA sequences may emerge due to the functional, structural or transformative relationship among them. DNA Sequence alignment is a strategy for arranging two (pairwise) or more (multiple) sequences of DNA to recognize locales of similarity among them via looking for a progression of individual nucleotides or nucleotide designs that are in the same order [23]. Aligned sequences represented as rows in a grid. Gaps ('-') need aid embedded between the characters with the goal. So that the same or similar characters are aligned in consecutive columns. Sequence alignment of two biological sequences may be called pairwise sequence alignment, also in the event more than two sequences are involved, it may be called multiple sequence alignment [24]. Here, in this paper, we use the pairwise sequence alignment in a global and local algorithms and examined the measure of the matching based on the scoring matrix for DNA alignment. And use the Neuro-Fuzzy model in the Matlab tool, that implemented by the data set files of measure score matching of DNA sequences that deal with the set of biological data. This tool is efficient and fast to evaluate the confidence scoring measure of matching the DNA sequences.

2. LITERATURE VIEW

The sequence alignment took a large space research in Bioinformatics field. The study of it has been growing and in persistent need, and the need for fast and efficient the algorithms increases. In the previous research work has been studied on providing new algorithms with the main purpose of the proposing the requirements of efficient sequence alignment, the techniques have been used all the latest with providing fast and efficient sequence alignment algorithms.

In [12] suggest a DNA sequence alignment, which uses quality information and a fuzzy inference implementation developed based on the features of DNA parts and a fuzzy logic system on improving methods with a DNA sequence alignment that uses DNA sequence quality information. In [21] prepare the novel single-GPU parallelizations of the Smith-Waterman algorithm implements for pairwise sequence alignment. In [11] suggest the fuzzy logic model for approximate matching of DNA subsequences. In [24] proposing a multiple sequence alignment algorithm which performs fuzzy logic to measure the similarity of sequences based on the fuzzy parameters. The algorithm is examined on few data sets of real biological sequences taken from NCBI bank and evaluating its efficiency through Sinic View tool.

In [1] suggest a new pattern matching technique defined as exact multiple patterns matching algorithms utilizes DNA sequence and pattern pair. The current method is used to avoid unneeded comparisons in the DNA sequence. In [15] explain the performing of the pairwise sequence alignments by the Biostrings bundle. In [16] uses the methods for inspection the protein and DNA evolution of protein and DNA sequence that have been working on the database that is Local comparison and Global Comparison.

In [17] performs the new representation in working out the similarity/ dissimilarities of the entire genomes and polynucleotides. In [18] utilizes the techniques for investigation the protein and DNA development of protein and DNA sequence that have been taking a work in the database that is the Local comparison and Global Comparison. In [19] DNA sequences develop by local changes influencing one or a few adjacent symbols, and in addition to large scale improvements and duplications. This outcome in mosaic sequences with different degrees of similarity between districts inside a solitary genome or in genomes of related life forms. In our work, we use the Neuro-Fuzzy model utilizes the biological dataset files for matching DNA, and measure the score of matching the DNA sequences with Global and Local alignment.

3. SEQUENCE ALIGNMENT

DNA sequence alignment is the Procedure for comparing two or more sequences by searching for a series of individual characters that are similar and in the same order in those sequences, sequence alignment techniques have two categories and is generally divided into :- Pairwise alignment: compare two sequences- Multiple sequence alignment: compare > 2 sequences. Here, we implement the pairwise method, the dynamic programming uses to align the sequences, to do the alignment method we do the three steps: Scoring Matrix, Backtrace, and the Alignment. The ways we use it to perform the alignment are: global and local alignment, these algorithms uses the proposed matrix to measure the similarity of bases in the two sequences, The best measuring score alignment is the alignment which ends in the cell in the matrix with the highest scoring value. Table 1 shows the matrix utilized within the proposed algorithms.

Table 1. The Matrix used within the proposed algorithms

S1 S2	A	C	G	T	Gaps
A	+1	-1	-1	-1	-2
C	-1	+1	-1	-1	-2
G	-1	-1	+1	-1	-2
T	-1	-1	-1	+1	-2
Gaps	-2	-2	-2	-2	-

3.1 Needleman–Wunsch algorithm

For the Needleman-Wunsch algorithm, a scoring matrix is. Ascertained for those two provided for sequences A and B, by setting one sequence along column side, furthermore on the turn sequence side. It is Additionally Frequently referred as optimal matching algorithm and the global alignment technique. Global alignment algorithms begin from the starting of two DNA sequences and put the gaps in each until reaching the end of one of both. Main develop a grid for example, such that those particular cases demonstrated in table 2.

Table 2. Matrix for Global alignment (GA)

S1 S2	-	A	T	G	T	C	C
-	0	-2	-4	-6	-8	-10	-12
A	-2	+1	-1	-3	-5	-7	-9
G	-4	-1	0	0	-2	-4	-6
T	-6	-3	0	-1	+1	-1	-3
C	-8	-5	-4	-1	-1	+2	0
C	-10	-7	-6	-3	-2	0	+3
A	-12	-9	-8	-5	-4	-2	**+1**

3.2 Smith Water Man Algorithm

The Smith–Waterman algorithm, which is the method used to perform the local sequence alignment, Local alignment algorithms find the sections of highest similarity between two sequences and create the alignment to abroad from there; that is, identify the most similar portion comparable sub-region imparted between two successions.

Table 3. Matrix for local alignment (LA)

S1 S2	-	T	C	A	G	T	T	G	C	C
-	0	0	0	0	0	0	0	0	0	0
A	0	0	0	1	0	0	0	0	0	0
G	0	0	0	0	2	0	0	1	0	0
G	0	0	0	0	1	0	0	1	0	0
T	0	1	0	0	0	2	1	0	0	0
T	0	1	0	0	0	1	3	1	0	0
G	0	0	0	0	1	0	1	4	2	0

The resulting alignment is: TCAGTTGC - AGGTTG the fundamental distinction with the Needleman-Wunsch algorithm is that negative scoring grid units need aid set to zero, which renders the (positive scoring) local alignments LA unmistakable. As shown in table 3 Backtracking begins during the most scoring grid cell Furthermore returns to the cell with score zero will be encountered, yielding the highest scoring local alignment.

4. THE PROPOSED METHOD

To implement our method in a sequence alignment, we use the Nero-Fuzzy technique in Matlab tool. When we complete the alignment method as in the table 2, and get the matching data about the (matching score, mismatching score and gaps), and calculate the resulting score according to the equation 5, we save this data in the dataset file to be training and testing on the ANFIS system.

The Neuro-fuzzy model is very well established approach and has a tremendous potentiality to outcome results with high accuracy ratio and the efficiency with biological data to determine the confidence measure score of matching DNA sequencing. Those recommended sequences-matching algorithm utilize the three input variables – match score (match), mismatch score (mismatch), and gaps, as shown in Fig. 1.

These three inputs would then fuzzified utilizing following membership functions equations and giving the calculated resulting score (score, computed utilizing a substitution matrix):

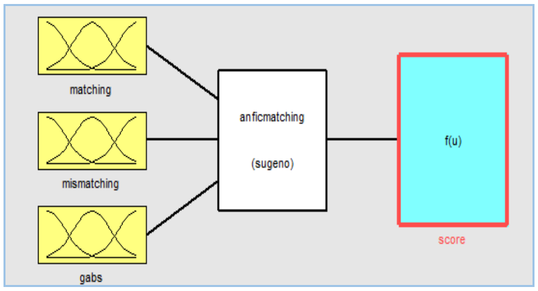


Figure 1: The three Input variables and the output

Then the description of each one will be as following:

$$\text{Matching=}\begin{cases} 0 \quad \text{if there is no similarity} & (1) \\ 1 \quad \text{if there is highest similarity 100\%} \\ [1,0] \quad \text{(matching score/enseq)} \end{cases}$$

$$\text{Mismatch=}\begin{cases} 0 \quad \text{if there is no mismatch} & (2) \\ 1 \quad \text{if there is no similarity} \\ [1,0] \quad \text{(mismatching score/ lenseq)} \end{cases}$$

$$\text{Gap=}\begin{cases} 0 \quad \text{if there is no need to put a gap} \\ [1,0] \quad \text{(gaps score/lenseq)} & (3) \end{cases}$$

$$\text{Score =}\begin{cases} 0 \quad \text{if the resulting score} \leq 0 & (4) \\ 1 \quad \text{if the resulting score = perfect score} \\ [1,0] \quad \text{(resulting score/perfect score)} \end{cases}$$

The score is measured in the ANFIS Tool in Matlab in this range, the variable" lenses" mean the entire length of the sequence. The score is measured according to the equation for the calculating the similarity of the pairwise sequencing:

Score = match + (- mismatch) + (- gabs) (5)

It is calculated from the three input value according to the substitution Matrix, see table.1.

5. THE SIMULATION RESULTS

In our work, we perform the Neuro-Fuzzy model by the ANFIS tool in Matlab, using the data set about 600 samples for a matching measure score of DNA sequencing, these data divide into two data files for the training and testing, for training step, we use the data set about 450 samples, and for testing step, we use the data about 150 samples. We get these data when we perform the pairwise sequence alignment, in the Global alignment GA we use the substitution matrix to get the optimal alignment, and from the equation 5. We calculate the confidence score matching. We use these dataset files in ANFIS system, and output the result in the range value in equation 1,2,3,4. We use different processing systems to implement the matching results, and each system has different results with convergent values, as shown in Fig. 2 Explain the membership function, the training and testing phase and the ANFIS structure for each, Table 4 shows the details of the various ANFIS testing results.

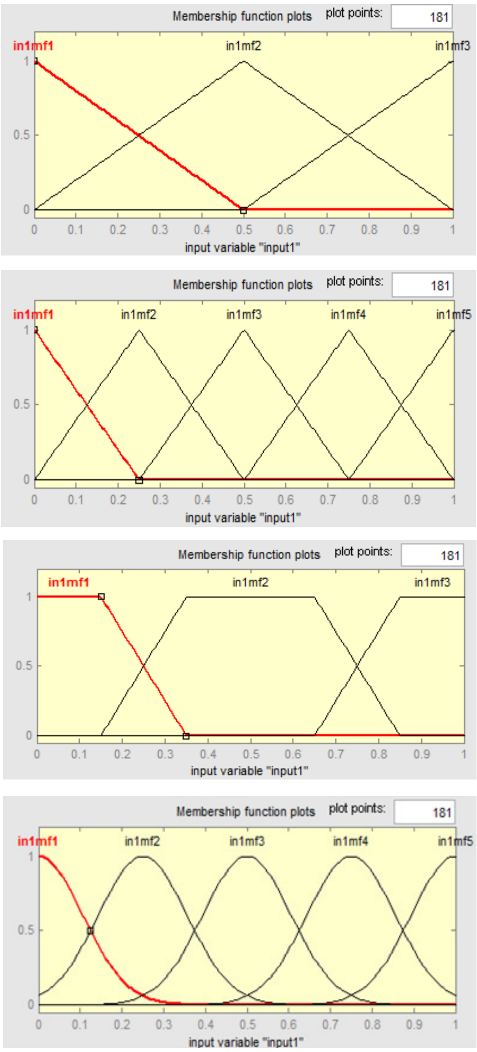


Figure 2: The different inputs of MF used with several processing of triangular, trapizoidal and Gaussian MFs

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4. The various ANFIS testing results

Different MFs	FIS method	The number Of epochs		Training error	The average testing error	
Trian. MF	3 3 3	Constant	Backpropagation	50	0.046574	**0.060781**
Trian. MF	5 5 5	Constant	Backpropagation	250	0.011085	**0.043061**
Trape .MF	3 3 3	Constant	Backpropagation	100	0.018984	**0.045302**
Trape .MF	3 3 3	Constant	Backpropagation	500	0.016572	**0.01657**
Gaus. MF	5 5 5	Constant	Backpropagation	350	0.014363	**0.044363**
Gaus. MF	5 5 5	Constant	Backpropagation	500	0.028041	**0.054149**

Here we will explain all the necessary models to check each state of results.

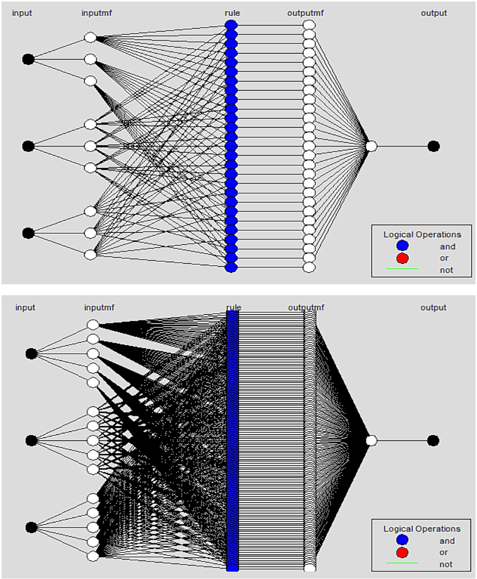
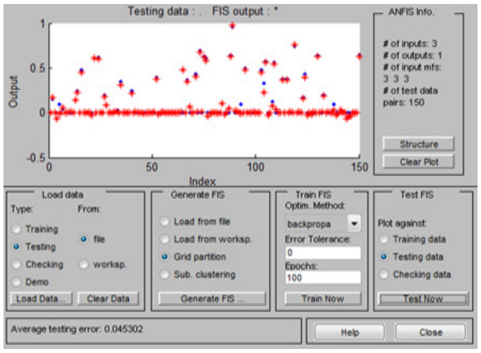


Figure 3: The ANFIS Structure used with several processing of ANFIS Structure with three MF ANFIS Structure with five MF


Testing data : . FIS output : *

ANFIS Info.
# of inputs: 3
# of outputs: 1
# of input mfs:
3 3 3
# of test data pairs: 150

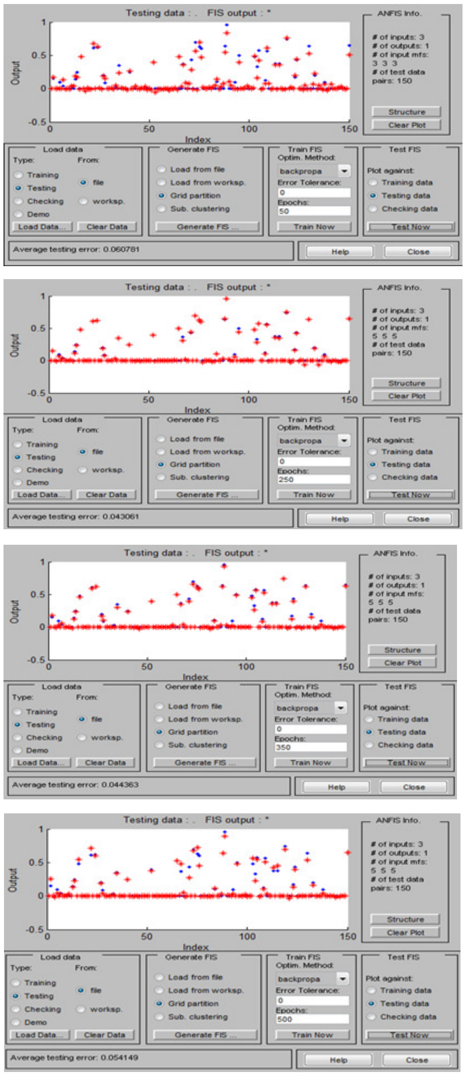
Load data
Type: ○ Training ● Testing ○ Checking ○ Demo
From: ● file ○ worksp.

Generate FIS
○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Train FIS
Optim. Method: backpropa
Error Tolerance: 0
Epochs: 100

Test FIS
Plot against: ○ Training data ● Testing data ○ Checking data

Average testing error: 0.045302


Testing data : . FIS output : *

ANFIS Info.
# of inputs: 3
# of outputs: 1
# of input mfs:
3 3 3
# of test data pairs: 150

Load data
Type: ○ Training ● Testing ○ Checking ○ Demo
From: ● file ○ worksp.

Generate FIS
○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Train FIS
Optim. Method: backpropa
Error Tolerance: 0
Epochs: 50

Test FIS
Plot against: ○ Training data ● Testing data ○ Checking data

Average testing error: 0.060781

Testing data : . FIS output : *

ANFIS Info.
# of inputs: 3
# of outputs: 1
# of input mfs:
5 5 5
# of test data pairs: 150

Load data
Type: ○ Training ● Testing ○ Checking ○ Demo
From: ● file ○ worksp.

Generate FIS
○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Train FIS
Optim. Method: backpropa
Error Tolerance: 0
Epochs: 250

Test FIS
Plot against: ○ Training data ● Testing data ○ Checking data

Average testing error: 0.043061

Testing data : . FIS output : *

ANFIS Info.
# of inputs: 3
# of outputs: 1
# of input mfs:
5 5 5
# of test data pairs: 150

Load data
Type: ○ Training ● Testing ○ Checking ○ Demo
From: ● file ○ worksp.

Generate FIS
○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Train FIS
Optim. Method: backpropa
Error Tolerance: 0
Epochs: 350

Test FIS
Plot against: ○ Training data ● Testing data ○ Checking data

Average testing error: 0.044363

Testing data : . FIS output : *

ANFIS Info.
# of inputs: 3
# of outputs: 1
# of input mfs:
5 5 5
# of test data pairs: 150

Load data
Type: ○ Training ● Testing ○ Checking ○ Demo
From: ● file ○ worksp.

Generate FIS
○ Load from file
○ Load from worksp.
● Grid partition
○ Sub. clustering

Train FIS
Optim. Method: backpropa
Error Tolerance: 0
Epochs: 500

Test FIS
Plot against: ○ Training data ● Testing data ○ Checking data

Average testing error: 0.054149

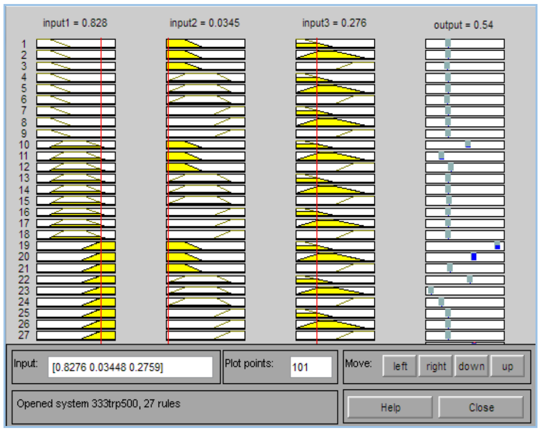
Figure 4: The different testing data used with several processing

In Figure 2,3,4 Explain the various ANFIS used with different MF, in these processing systems gets different results, through the training and testing phase, there is different average training and testing errors, we choose the most suitable ANFIS system with lowest average testing error, in table 4 shows the different processing system we implement it, with the details.

However, in this table the most suitable system is the Trapizoidal MF ANFIS, because it has the lowest average testing error, and give the result with high performance, thus we use it to get the confidence score matching of the numeric data from the DNA sequence alignment. Show table 5 and table 6.

Sequence alignment is a necessary condition for analysing DNA sequencing, in this method we use the numeric biological data for sequence alignment, and we use it as the input in our model. The ANFIS system is appropriate to our data, in this system implement several processing systems to get the most suitable results as shown in table 4, the most suitable system is the trapezoidal with three membership function for each input and 500 epochs, it has the lowest average testing error. In our method we use the pairwise sequence alignment, it is applied using the Global and Local alignment algorithm method as shown in table 5,6.

We use several different sequences to be aligned, as shown in table 5, we aligned the sequences in the global alignment algorithm, in this method insert the gaps when the base in the sequence not similar with the other in the same order in this pair, as the way in table 2, we shifted the character and input the gap in order to similar the character with the other character in the same order in this pair, we calculate the number of times for similar character (matching) and the number of times for mismatching character (mismatching) and the number of times for gaps, and use it as the input in the ANFIS view tool, and output the resulting score matching, as shown in figure 5, we can compute the percentage similarity of the alignment using the view alignment in matlab, as show in figure 6. This use the number of times similar in the sequence alignment divided on the entire length of the alignment sequence. This method is used by Neuro-Fuzzy technique to get the better results with high accuracy, this technique is very efficient utilize with the biological data this is achieving the high performance with the previous work implemented in this topic.

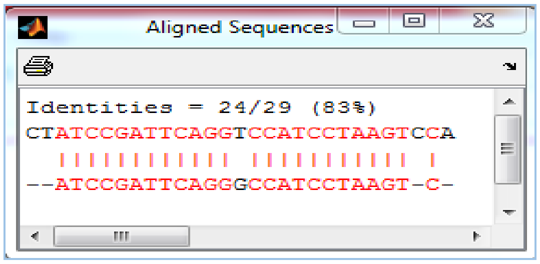


[0.8276 0.03448 0.2759]

101

Opened system 333trp500, 27 rules

Figure 5: View rules ANFIS



```
Identities = 24/29 (83%)
CTATCCGATTCAGGTCCATCCTAAGTCCA
  |||||||||||| ||||||||||| |
--ATCCGATTCAGGGCCATCCTAAGT-C-
```

Figure 6: The identical alignment

In table 6, we aligned the sequence using the local algorithm, in this method the algorithm take the most similar part of the pair sequence, not must in the order and not need to input the gap, the resulting score is the perfect there is no mismatch and no gaps, and have 100% identical.

6. CONCLUSION

The proposed technique here, is used to obtain the similarity measure of the pairwise DNA sequence alignment, the Pattern matching is an essential task of example the disclosure process in this day and age for finding the basic and utilitarian conduct in the DNA sequencing. In spite of the fact that example of matching is commonly is utilized as a part of computer science and information processing.

In this paper used the substitution matrix within the proposed algorithms in the sequence alignment, the proposed algorithms are the Global alignment and the Local alignment using to measure the score matching, which are utilized by the substitution matrix, and utilized it as a method to be align the two DNA sequences, and take the information from the resulting alignment which are the (match, mismatch and gaps) and use it as the input in our model.

The Neuro-Fuzzy model is used to evaluate the confidence measure score similarity by the ANFIS tool in Matlab, we implement the method in several processing systems, and depend on the most suitable system with lowest average testing error, we obtain the score matching result for several patterns of DNA sequencing, this model presented the matching implementation in fast and efficient.

REFERENCES

[1] Raju Bhukya and DVLN Somayajulu. Exact Multiple Pattern Matching Algorithm using DNA Sequence and Pattern Pair. International Journal of Computer Applications. 2011; (0975 – 8887).

[2] Xiaojing Xie, J. Guan, S.Zhou1. Similarity evaluation of DNA sequences based on frequent patterns and entropy. International Symposium on Bioinformatics Research and Applications (ISBRA-14), Zhangjiajie, China. June 2014; 28-30.

[3] Christopher J. O. Baker, Greg Butler, Volker Haarslev. Ontologies, semantic web and intelligent systems for genomics. 2014.

[4] Raed, I.H., Ahson, S.I. Confidence value prediction of DNA sequencing with Petri net model Journal of King Saud University – Computer and Information Sciences. 2011; 23: 79–89.

[5] Raed I. Hamed, S. I. Ahson, R. Parveen. A New Approach for Modelling Gene Regulatory Networks Using Fuzzy Petri Nets. Journal of Integrative Bioinformatics. 2010; 7(1):113.

[6] Hong-Jie Yu. Similarity Analysis of DNA Sequences Based on Three 2-D Cumulative Ratio Curves. ChapterBio-Inspired Computing and ApplicationsVolume 6840 of the series Lecture Notes in Computer Science. 2012; pp 462-469.

[7] Lukas Habegger. Computational Methodologies for Transcript Analysis in the Age of Next-Generation DNA Sequencing. 2012.

[8] Wei Deng1, 2 and Yihui Luan1. Analysis of Similarity/Dissimilarity of DNA Sequences Based on Chaos Game Representation. Hindawi Publishing Corporation. 2013.

[9] Pandiselvam.P, Marimuthu.T, Lawrance. R. A COMPARATIVE STUDY ON STRING MATCHING ALGORITHMS OF BIOLOGICAL SEQUENCES. 2014.

[10] Martina Višˇnovská, Tomáš Vinaˇr, Broˇna Brejová. DNA Sequence Segmentation Based on Local Similarity. ITAT Proceedings, CEUR Workshop Proceedings. 2013; Vol. 1003, pp. 36–43.

[11] Hamed, R.I., Ahson, S.I., Parveen, R.. Designing genetic regulatory networks using fuzzy Petri nets approach. Int. J. Autom. Comput. 7, 403–412, 2010.

[12] Kwangbaek Kim, Minhwan Kim, Youngwoon Woo. A DNA sequence alignment algorithm using quality information and a fuzzy inference method. Science direct, Progress in Natural Science. 2008; 18: 595–602.

[13] S. B. Needleman and C. D.Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Molecular Biolog. 1970; 48: 443-453.

[14] T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. Molecular Biology. 1981; 147: 195-197.

[15] Patrick Aboyou. Pairwise Sequence Alignments. Gentleman LabFred Hutchinson Cancer Research Center, Seattle, WA. 2016.

[16] Sudha M.P, P.Sripriya. Sequence Alignment in DNA Using Smith Waterman and Needleman Algorithms. (IJCSIT) Internat. Journal of Computer Science and Information Technologies. 2014.

[17] Subhram Das, Debanjan De, D. K. Bhattacharya. Similarity and dissimilarity of whole genomes using intuitionistic fuzzy logic. Notes on Intuitionist. Fuzzy Sets. 2015.

[18] M.P Sudha, P. Sripriya. Sequence Alignment in DNA Using Smith Waterman and Needleman Algorithms. (IJCSIT) Internati. Journal of Computer Science and Information Technologies. 2014.

[19] Martina Višˇnovská, Tomáš Vinaˇr, Broˇna Brejová. DNA Sequence Segmentation Based on Local Similarity. ITAT Proceedings, CEUR Workshop Proceedings. 2013; Vol. 1003, pp. 36–43.

[20] Mark Craven, Pairwise Sequence Alignment. (Part 1), 2002; BMI/CS 576.

[21] Junjie Li, Sanjay Ranka, and Sartaj Sahni. Pairwise Sequence Alignment for Very Long Sequences on GPUs. 2012.

[22] Ana Abecasis, Anne-Mieke Vandamme, Philippe Lemey. Reviews 2 Sequence Alignment Sequence Alignment in HIV Computational Analysis. 2008.

[23] Tamal Chakrabarti, Sourav Saha, Devadatta Sinha. DNA Multiple Sequence Alignment by a Hidden Markov Model and Fuzzy Levenshtein Distance based Genetic Algorithm. International Journal of Computer Applications. July 2013; (0975 – 8887) Volume 73– No.1.

[24] Nivit Gill and Shailendra Singh. Biological Sequence Matching Using Fuzzy Logic. International Journal of Scientific & Engineering Research. 2011.

Table 5. The score matching for Global Alignment of ANFIS Tool

Sequences	Global alignment	Identities %	Score
AGTAAGTTAGCAAGTAGTTCAGTCCTGC CGCCCGGCTCCGTT	AGTAAGTTAGCAAGTAGTTCAGTCCTGCCGCCCGG - - - - - - - - - - - -- - - - - - - - - - - - - - - - - CT - CCG - TT - -	5/35 (14%)	**0**
AGCCGAGAATAATTAAGCCGTCCA ATGTCC	AGCCGAGAATAATTAAGCCGTCCA - - - - - - - - - - - - A - T - - - - - GTCC -	6/24 25%)(	**0**
AGGTTGCAGGTC	AGGTTGC AGGT-- C	5/7 (71%)	**0.149**
GTAGGCTTAAGGTTATAGATC	GTAGGCTTAAGGTTA - T AG - - - - -A- - - T -C	5/15 (33%)	**0**
CTATCCGATTCAGGTCCATCCTAAGTCC A ATCCGATTCAGGGCCATCCTAAGTC	CTATCCGATTCAGGTCCATCCTAAGTCCA - - ATCCGATTCCGGGCCATCCTAAGT-C-	24/29 (83%)	**0.54**
AGTCCAATGTCC	A - GTCCA A TGTCC -	5/7 (71%)	**0.149**
ACCATGATTCCATTCGTATTCTAATACCG GCAATAACATTCGGACTTACGTC	ACCATGATTCCATTCGTATTCTAATACCGGCAAT - - -A - - A - - -CATTCGGA- - CT- - TA –CGTC- - -	16/34 47%	**0**
CGGGAATTGAC	CGGGA- ATTGAC	2/6 33%	**0**
CTATCCGCTAGTCG	CTATCCG CTAGTCG	5/7 (71%)	**0.425**

Table 6. The score matching for Local Alignment of ANFIS Tool

Sequences	Local alignment	Identities%	Score
AGTAAGTTAGCAAGTAGTTCAGTCCTGCCGCCCGG CTCCGTT	CCG CCG	100%	**1**
AGCCGAGAATAATTAAGCCGTCCA ATGTCC	GTCC GTCC	100%	**1**
AGGTTGC AGGTC	AGGT AGGT	100%	**1**
GTAGGCTTAAGGTTA TAGATC	TAG TAG	100%	**1**
AGTCCA ATGTCC	GTCC GTCC	100%	**1**
ACCATGATTCCATTCGTATTCTAATACCGGCAAT AACATTCGGACTTACGTC	CATTCG CATTCG	100%	**1**
CGGGA ATTGAC	GA GA	100%	**1**
CTATCCG CTAGTCG	CTA CTA	100%	**1**